

Эмпирическое моделирование и прогноз эволюции открытых динамических систем: общий подход и примеры из климата

Александр Фейгин

*Федеральный исследовательский центр
«Институт прикладной физики РАН»*

СОАВТОРЫ:

*Дмитрий Мухин, Андрей Гаврилов, Алексей Селезнев,
Евгений Лоскутов*

*Федеральный исследовательский центр
«Институт прикладной физики РАН»*

Эмпирическая модель:

- **Оператор эволюции** (модель, описывающая изменение состояние системы с течением времени), **построенный путем прямого анализа данных**

**Эмпирическое моделирование
востребовано, если модель «из первых
принципов» отсутствует либо ее динамика
недостаточно адекватна имеющимся
данным наблюдений**

Объекты эмпирического моделирования:

сложные

*(нелинейные, высокоразмерные, мультимасштабные,
пространственно распределенные)*

динамические системы,

подверженные внешним воздействиям различного типа

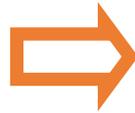
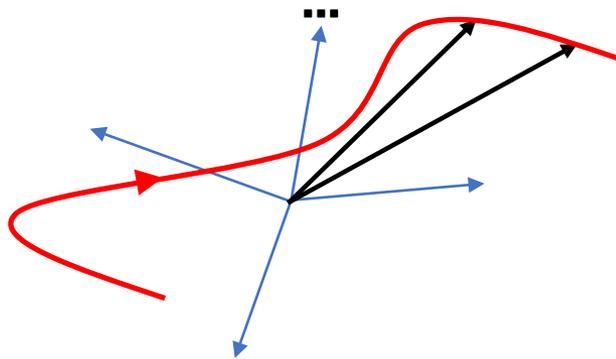
Примеры сложных систем:

- **климатические системы;**
 - **живые системы;**
- **социальные системы;**
 - ...

Ситуация, идеальная для эмпирического моделирования:



В фазовом пространстве размерности d строим фазовую траекторию и решаем задачу аппроксимации:



аппроксиматор
(оператор эволюции)

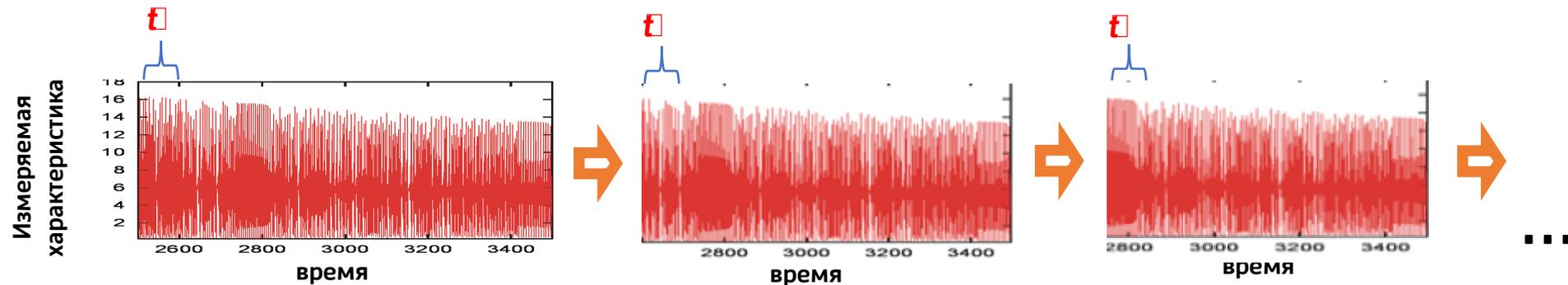
параметры
аппроксимации

Реальность в случае сложных природных систем:

(1) измерены временные ряды малого числа (одной!) переменных !?

Что можно сделать:

*использовать в качестве фазовых координат
«переменные с задержкой» (Takens, 1980)*



Реальность в случае сложных природных систем:

**(2) протяженность временных
рядов конечна**



ОГРАНИЧЕНИЕ МАКСИМАЛЬНОЙ
РАЗМЕРНОСТИ РЕКОНСТРУИРУЕМОГО
ФАЗОВОГО ПРОСТРАНСТВА:

$$d_{max} \sim \log\{\text{длина временного ряда}\}$$



СТОХАСТИЗАЦИЯ РЕКОНСТРУИРУЕМОЙ
ФАЗОВОЙ ТРАЕКТОРИИ

**(3) временное разрешение
конечно**



«БЫСТРАЯ» ЭВОЛЮЦИИ СИСТЕМЫ
ПРИСУТСТВУЕТ В ДАННЫХ В ФОРМЕ
СТОХАСТИЧЕСКОГО ПРОЦЕССА



СТОХАСТИЗАЦИЯ РЕКОНСТРУИРУЕМОЙ
ФАЗОВОЙ ТРАЕКТОРИИ

Что можно сделать:

Строить модель в форме низкоразмерной **стохастической**
динамической системы (*random dynamical system*)

АЛГОРИТМ

ЭМПИРИЧЕСКАЯ МОДЕЛЬ В ФОРМЕ СЛУЧАЙНОЙ ДИНАМИЧЕСКОЙ СИСТЕМЫ:

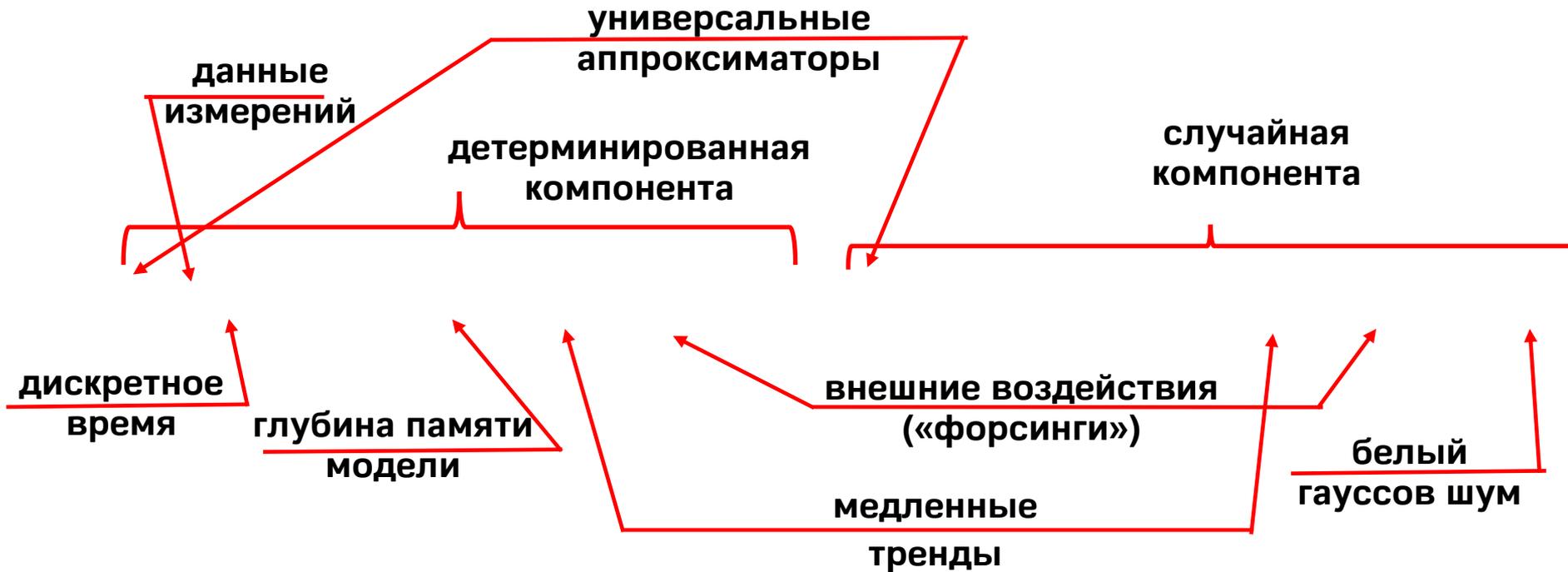
$$U_{n+1} = (U_n, \eta),$$

η - **неизвестный** случайный процесс



АЛГОРИТМ

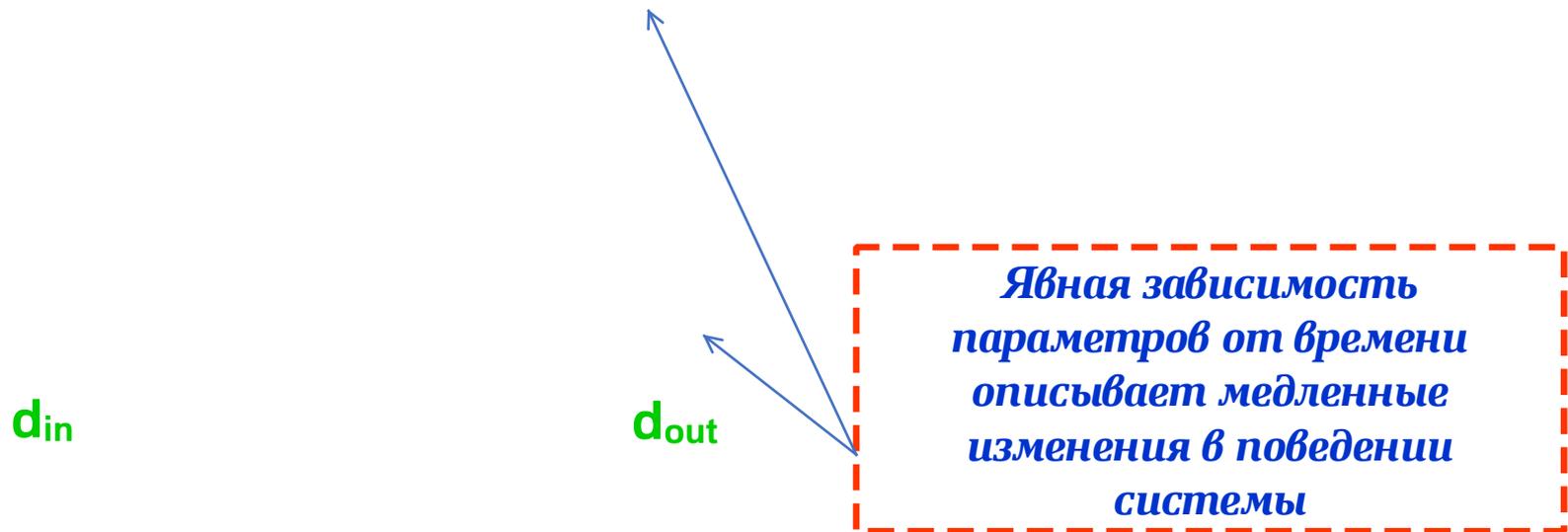
ЭМПИРИЧЕСКАЯ МОДЕЛЬ В ФОРМЕ СЛУЧАЙНОЙ ДИНАМИЧЕСКОЙ СИСТЕМЫ: УЧЕТ ДОПОЛНИТЕЛЬНЫХ ФАКТОРОВ



L – глубина памяти модели
 (t) – медленные тренды параметров
 $q(t)$ – внешние воздействия

ПРИМЕР УНИВЕРСАЛЬНОГО АППРОКСИМАТОРА:

Параметризация f и g искусственной нейронной сетью (ANN)



Два сорта параметров:

«Структурные» или «гипер» параметры
(«параметры сложности»):

Параметры модели заданной
сложности:

АЛГОРИТМ: ЕЩЕ ДВЕ ПРОБЛЕМЫ

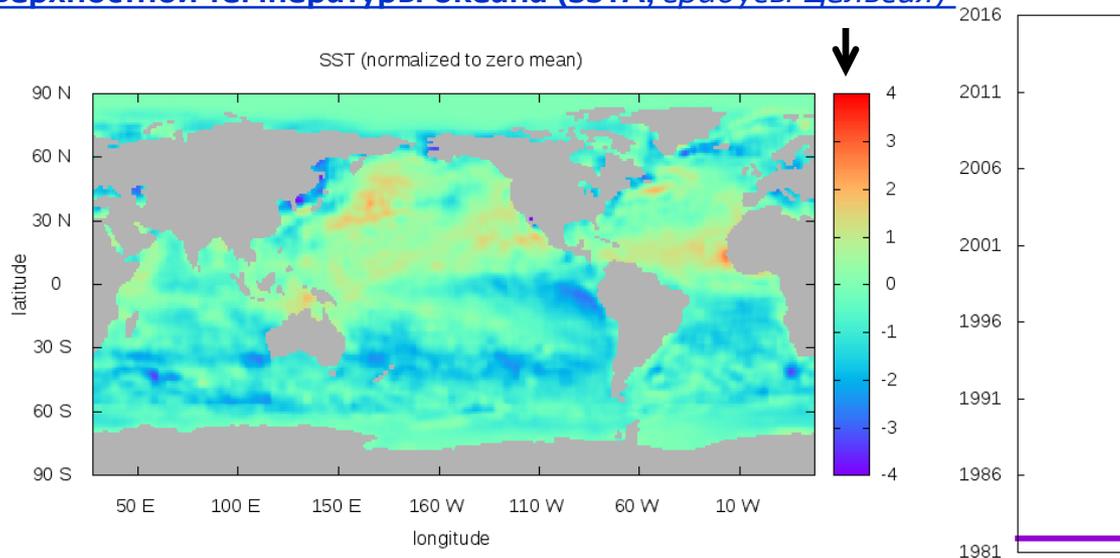
(4) **Высокая размерность данных**

ПРИМЕР: исследование современной климатической системы

Спутниковые данные: продолжительность \square 35 years (число отсчетов $N=396$)

http://iridl.ldeo.columbia.edu/SOURCES/.IGOSS/.nmc/.Reyn_SmithOIv2/.monthly/.sst/

Аномалии поверхностной температуры океана (SSTA, градусы Цельсия) t (годы)



Размерность данных (число временных рядов, измеренных в различных пространственных точках): $D=44219$

Максимальная размерность эмпирической модели $d_{max} \square \square \log N = 3!$



НЕОБХОДИМА ДЕКОМЗИЦИЯ (СНИЖЕНИЕ РАЗМЕРНОСТИ) ДАННЫХ

АЛГОРИТМ: ЕЩЕ ДВЕ ПРОБЛЕМЫ

*(5) Множественность решений задач реконструкции оператора эволюции и декомпозиции данных:
обе задачи являются некорректными обратными*



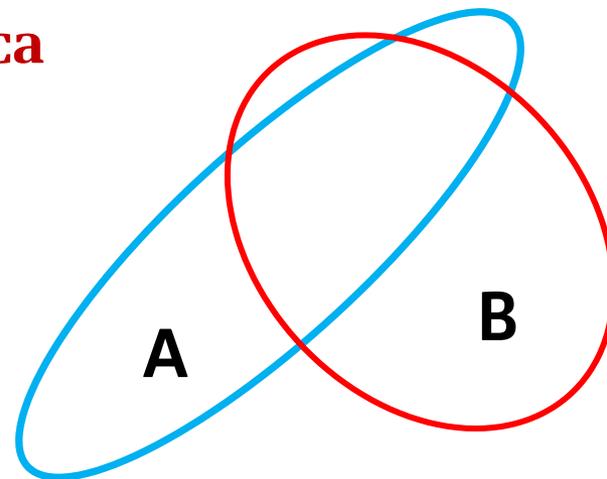
Необходимы адекватные исследуемой системе
алгоритмы оптимизации и обучения
как при построении модели, так и при декомпозиции данных



БАЙЕСОВ ПОДХОД К РЕШЕНИЮ ОБРАТНЫХ ЗАДАЧ

БАЙЕСОВ ПОДХОД К РЕШЕНИЮ ОБРАТНЫХ ЗАДАЧ

Теорема Байеса



ПОИСК ОПТИМАЛЬНЫХ ПАРАМЕТРОВ МОДЕЛИ

Замечание про разделение параметров на две группы:

«Структурные» параметры («параметры сложности») и
параметры модели заданной сложности

ОБУЧЕНИЕ МОДЕЛИ:

ОТЫСКИВАНИЕ ОПТИМАЛЬНОЙ МОДЕЛИ ОПТИМАЛЬНОЙ СЛОЖНОСТИ

1. Ценовая функция для поиска

оптимальных параметров модели заданной сложности:

ПВ параметров модели μ при условии данных измерений
и сложности модели S

Модель:

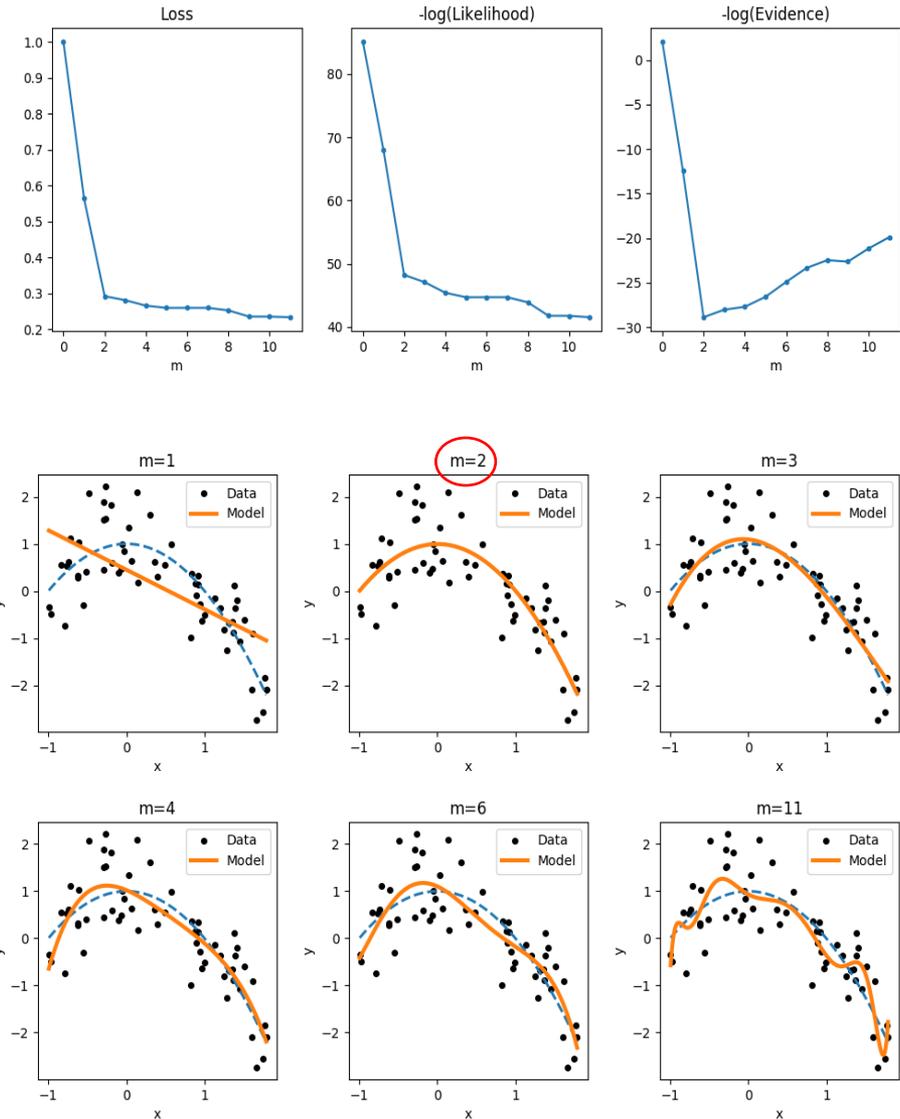


2. Ценовая функция для сложности модели: байесова обоснованность

Максимум $E(s)$ определяет оптимальное соотношение между качеством воспроизведения данных и прогностической способностью модели

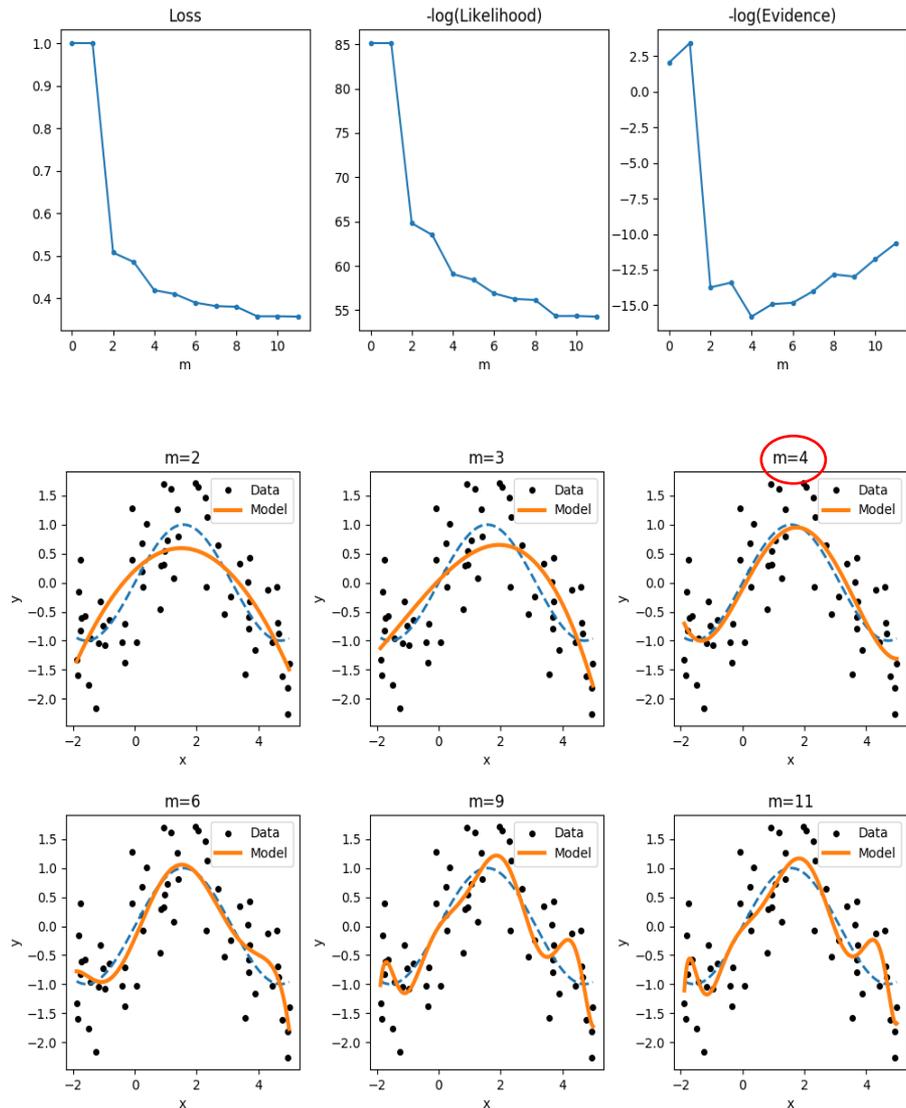
Байесов подход к решению обратных задач

Байесов критерий выбора модели



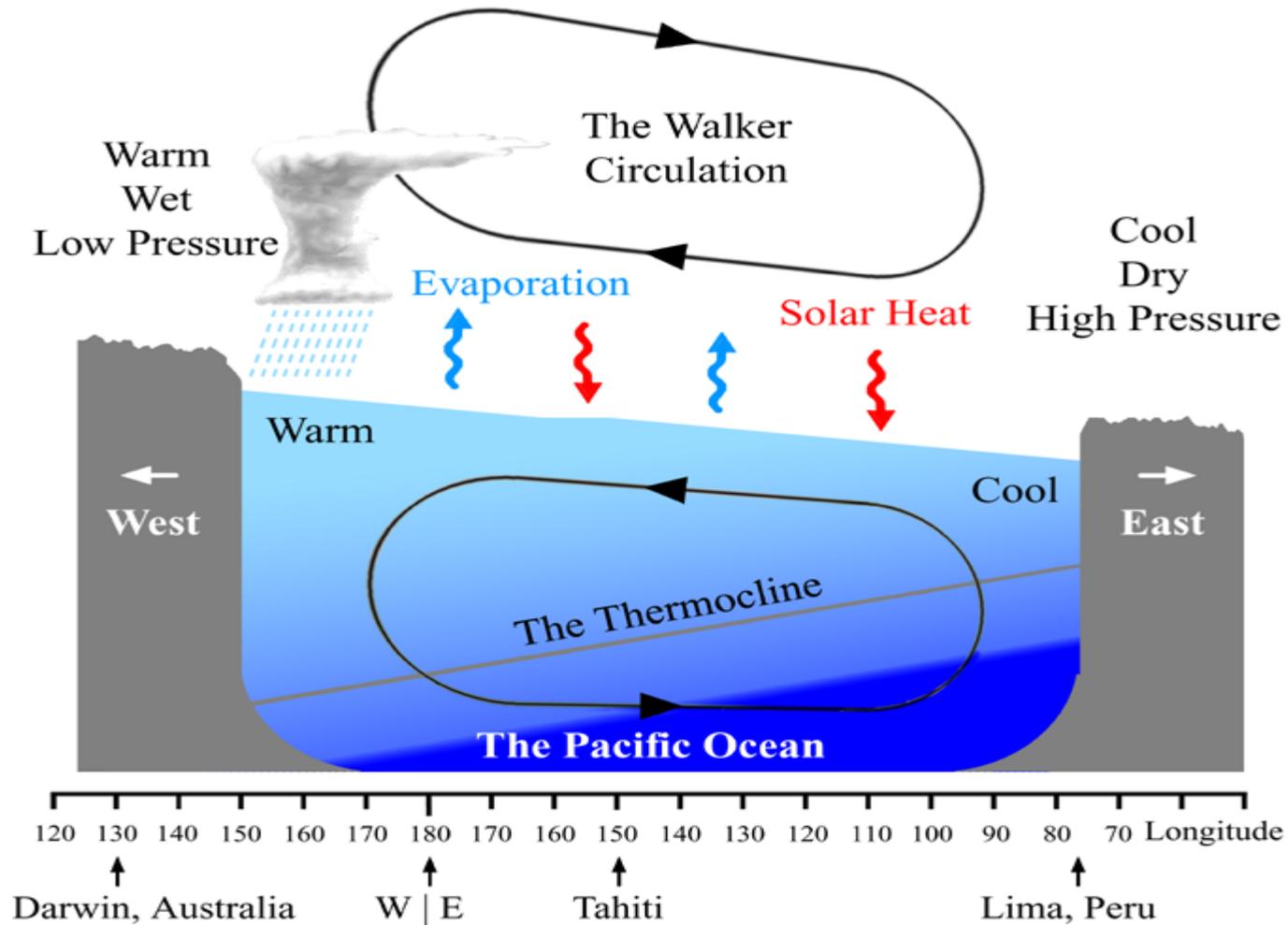
Байесов подход к решению обратных задач

Байесов критерий выбора модели



ЭМПИРИЧЕСКАЯ МОДЕЛЬ ЯВЛЕНИЯ ЭЛЬ-НИНЬО

СХЕМА НЕЙТРАЛЬНОЙ ФАЗЫ ЭЛЬ-НИНЬО - ЮЖНОГО КОЛЕБАНИЯ

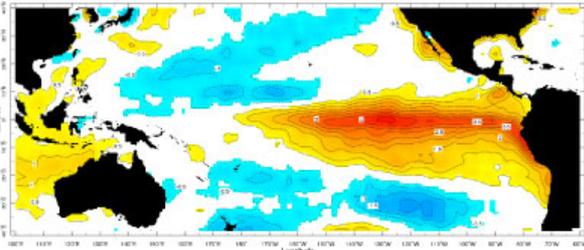


Уровень моря примерно на 60 сантиметров, а температура его поверхности примерно на 8–10 ° C выше в западной части Тихого океана, чем в восточной части Тихого океана.

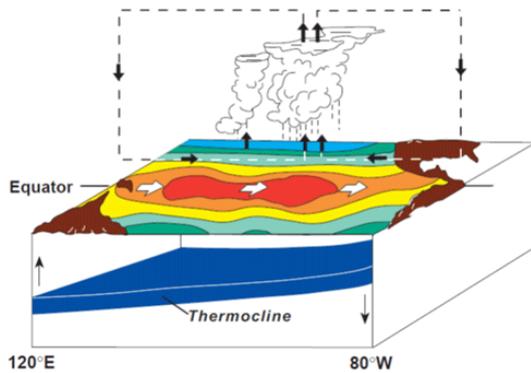
ТРИ ФАЗЫ ЭЛЬ-НИНЬО - ЮЖНОГО КОЛЕБАНИЯ

Эль-Ниньо

El Niño Episode Sea Surface Temperatures
Departure from average in degrees Celsius
Dec 1982 - Feb 1983



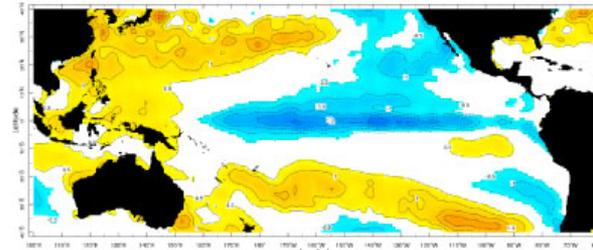
El Niño Conditions



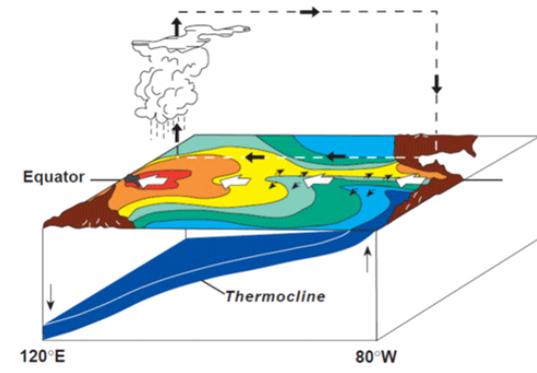
- Сильные аномалии (отклонения от климатического среднего) поверхностной температуры воды в центральной и восточной части тропического Тихого океана
- Сильные события случаются каждые 3-8 лет и продолжаются в течение одного года
- Пик аномалий поверхностной температуры, как правило, приходится на зимние месяцы

Ла-Нинья

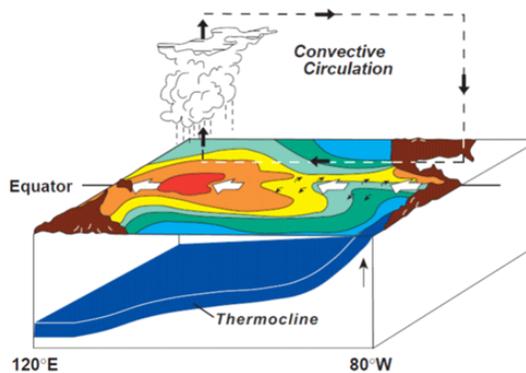
La Niña Episode Sea Surface Temperatures
Departure from average in degrees Celsius
Dec 1998 - Feb 1999



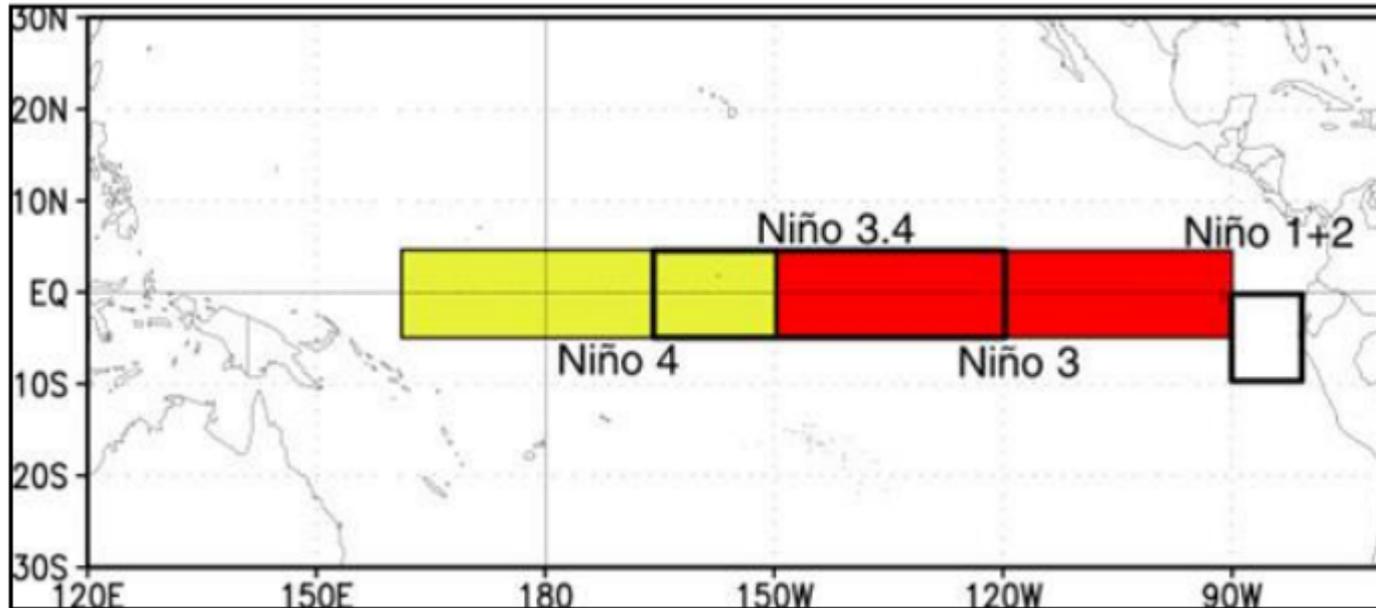
La Niña Conditions



Normal Conditions



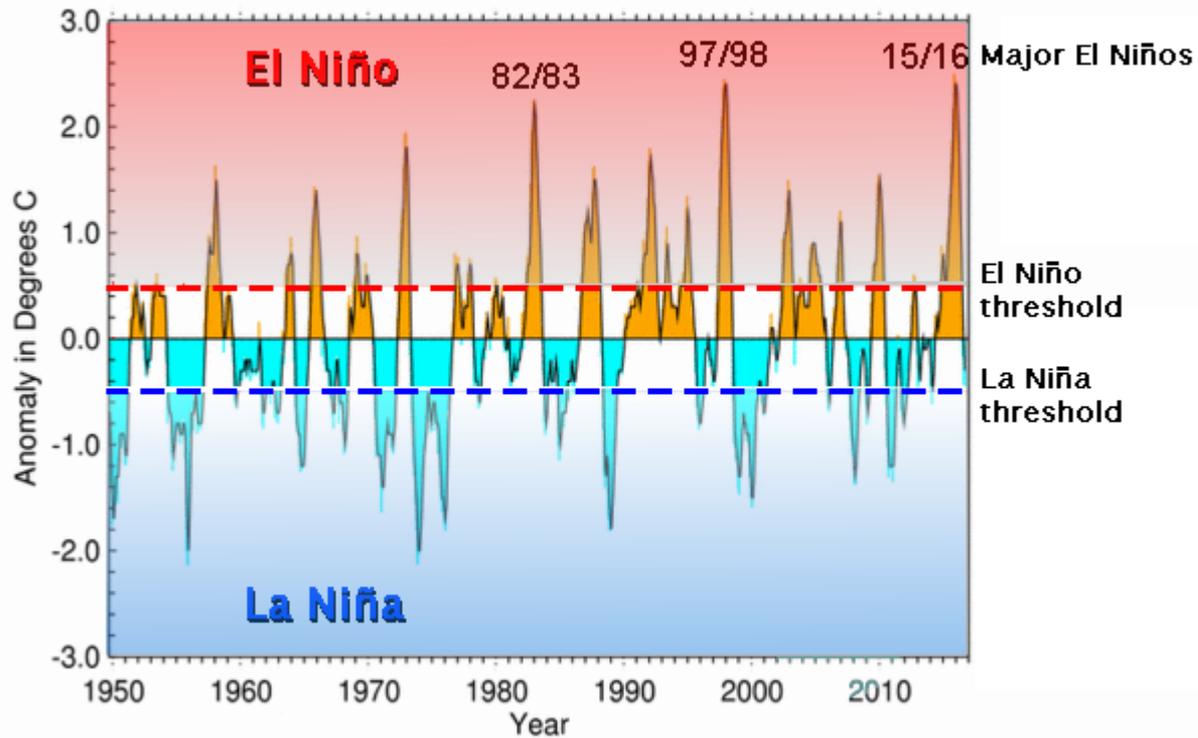
ИНДЕКСЫ ЭЛЬ-НИНЬО



Регионы экваториальной части Тихого океана, средняя аномалия температуры поверхности моря в которых используется для определения текущей фазы ЭНЮК (нейтральной, теплой или холодной).

ИНДЕКС Nino 3.4

SST Anomaly in Nino 3.4 Region (5N-5S,120-170W)

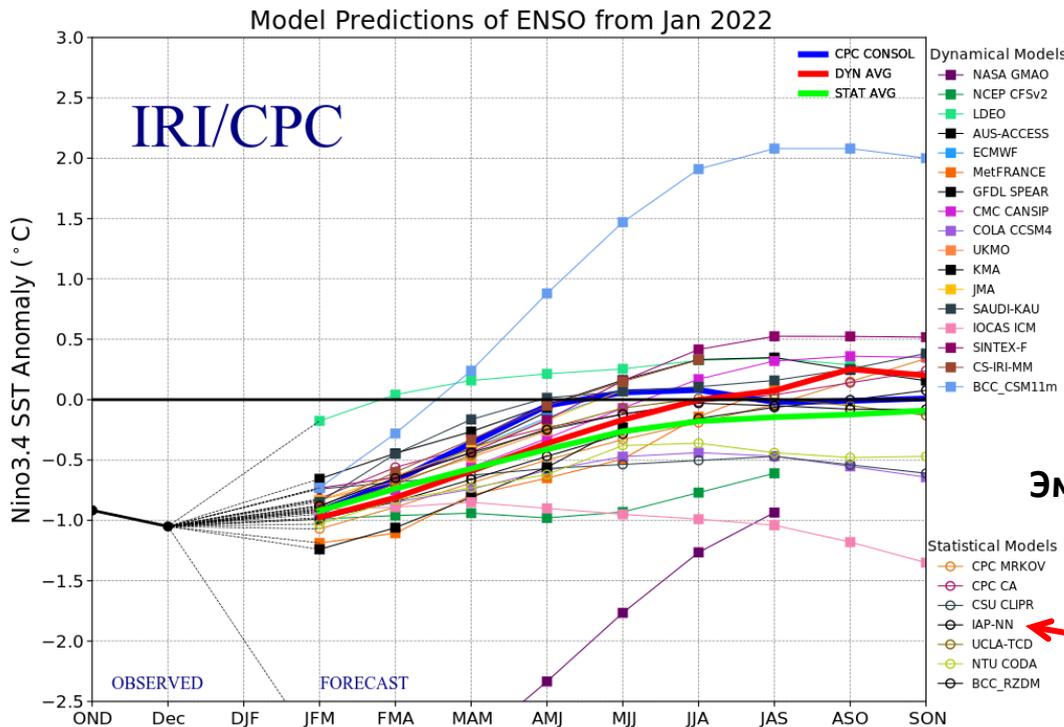


National Centers for Environmental Information / NESDIS / NOAA

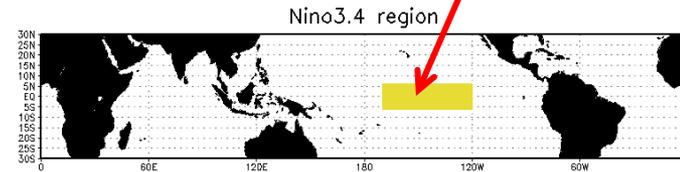
Внутригодовой прогноз изменчивости колебания Эль-Ниньо

https://iri.columbia.edu/our-expertise/climate/forecasts/enso/current/?enso_tab=enso-sst_table

Динамические модели



Индекс Nino3.4 - аномалии поверхностной температуры, усредненные в регионе 5N-5S, 170W-120W



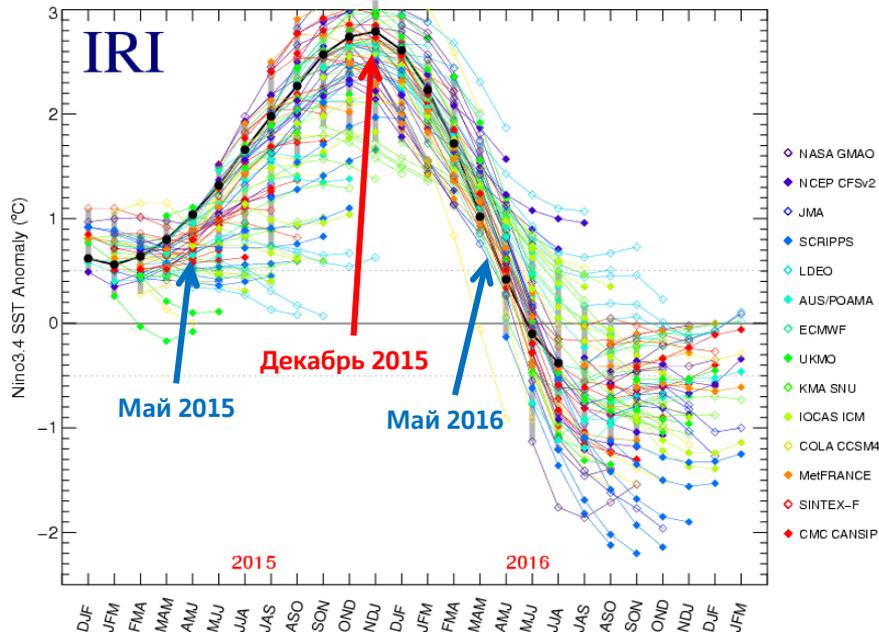
Эмпирические модели

Модель ИПФ РАН

Проблема весеннего барьера при прогнозе изменчивости колебания Эль-Ниньо

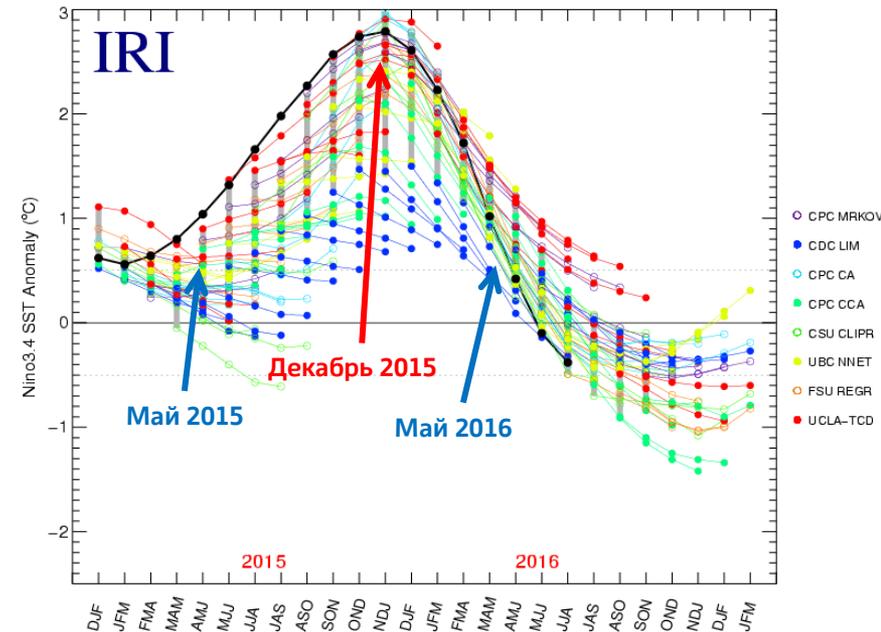
Динамические модели

ENSO Predictions for dynamical models, Dec 14 – Sep 16



Эмпирические модели

ENSO Predictions for statistical models, Dec 14 – Sep 16



Качество внутригодового прогноза индексов Эль-Ниньо с помощью современных моделей зависит от месяца, с которого прогноз стартует, и существенно хуже для зимних и весенних (до апреля-мая) стартовых месяцев

1. Предиктор внутригодовой изменчивости явления Эль-Ниньо на основе данных теплосодержания верхнего слоя тропического Тихого океана

Seleznev Aleksei, Mukhin Dmitry. *Improving statistical prediction and revealing nonlinearity of ENSO using observations of ocean heat content in the tropical Pacific // Climate Dynamics. 2021, no. 0123456789.*

Предобработка (декомпозиция) данных:

ПРОЕКЦИЯ НА БАЗИС ЭМПИРИЧЕСКИХ ОРТОГОНАЛЬНЫХ ФУНКЦИЙ

We have climate space-time field:

(n is time, N is time series duration, D is number of grid points)

Data rotation to EOF basis:

- new variable vector,

- orthogonal matrix

where $d, d \ll D$ of leading EOFs accounts for a major fraction of climate variance.



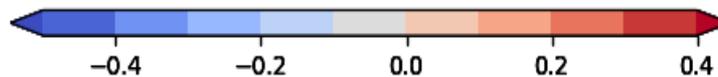
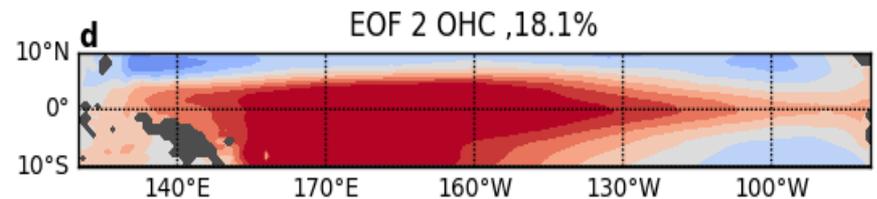
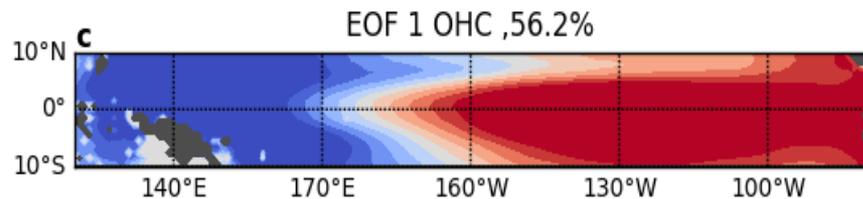
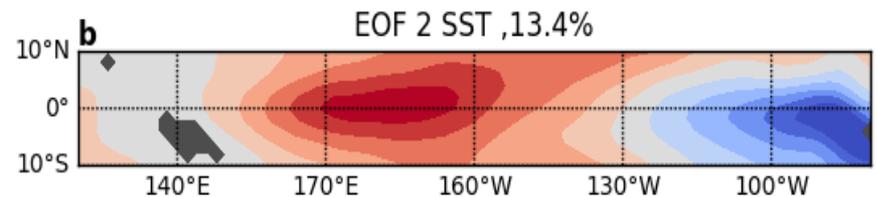
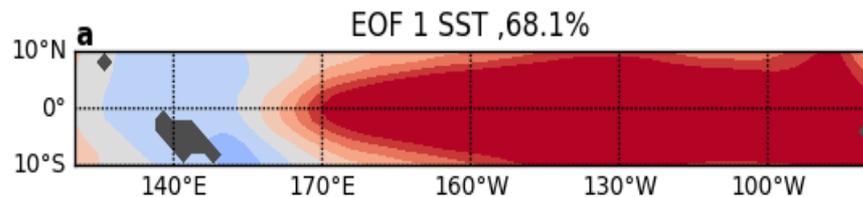
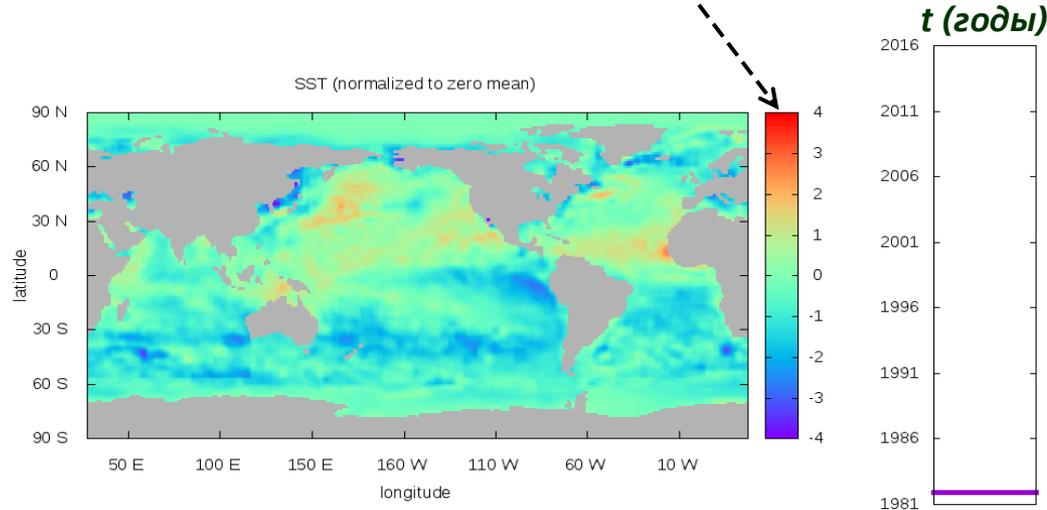
Covariation matrix of data

- orthogonally in time,

- orthogonally in space

Декомпозиция данных ТПО и ТО:

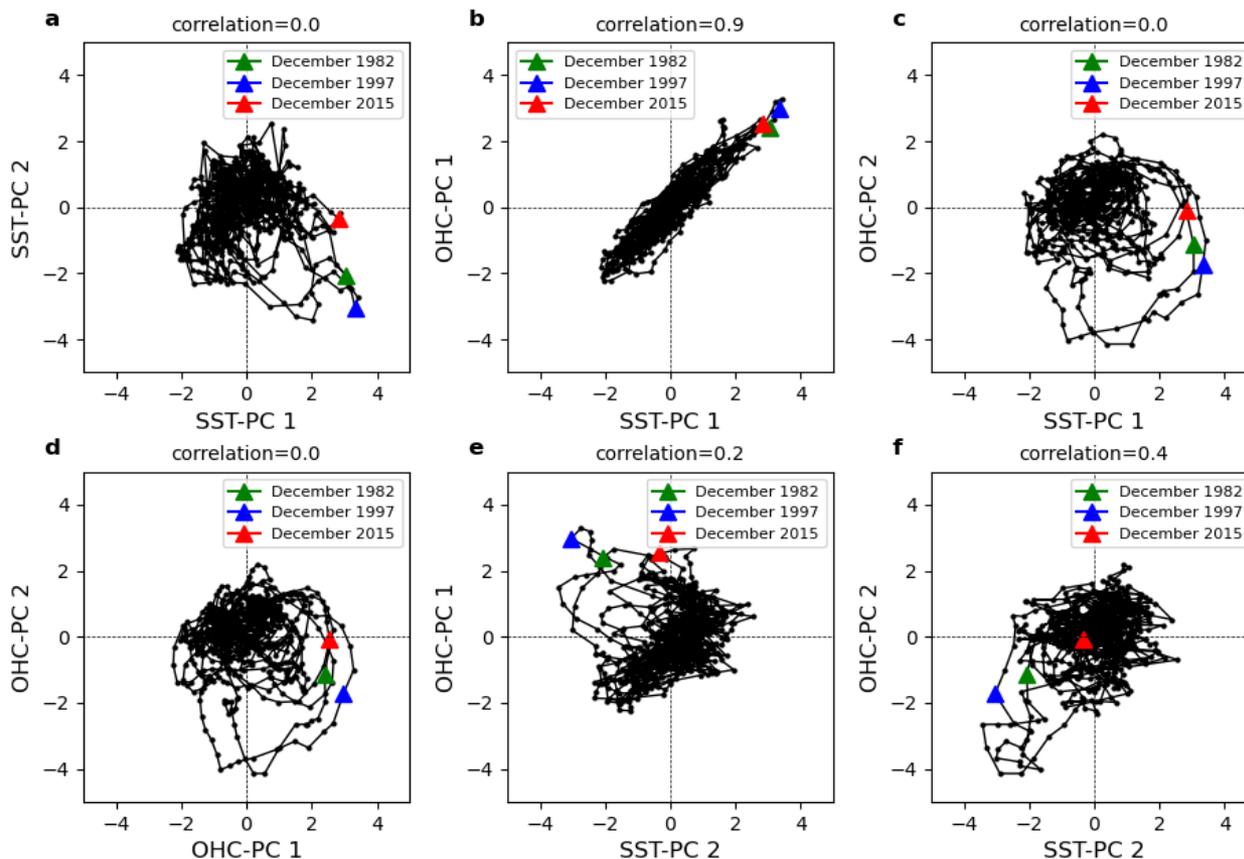
Аномалии поверхностной температуры океана (SSTA, градусы Цельсия) в тропической части Тихого океана (10° с.ш. - 10° ю.ш., 80° з.д. - 120° в.д.) с января 1960 по декабрь 2020 года



Пространственные распределения, соответствующие **двум главным ЭОФ аномалий ТПО** (верхняя панель, цветовая шкала в °C) и **ТО** (нижняя панель, цветовая шкала в $10^9 \text{J}/\text{m}^2$).

Доли захваченной вариации в процентах показаны сверху для каждой ЭОФ

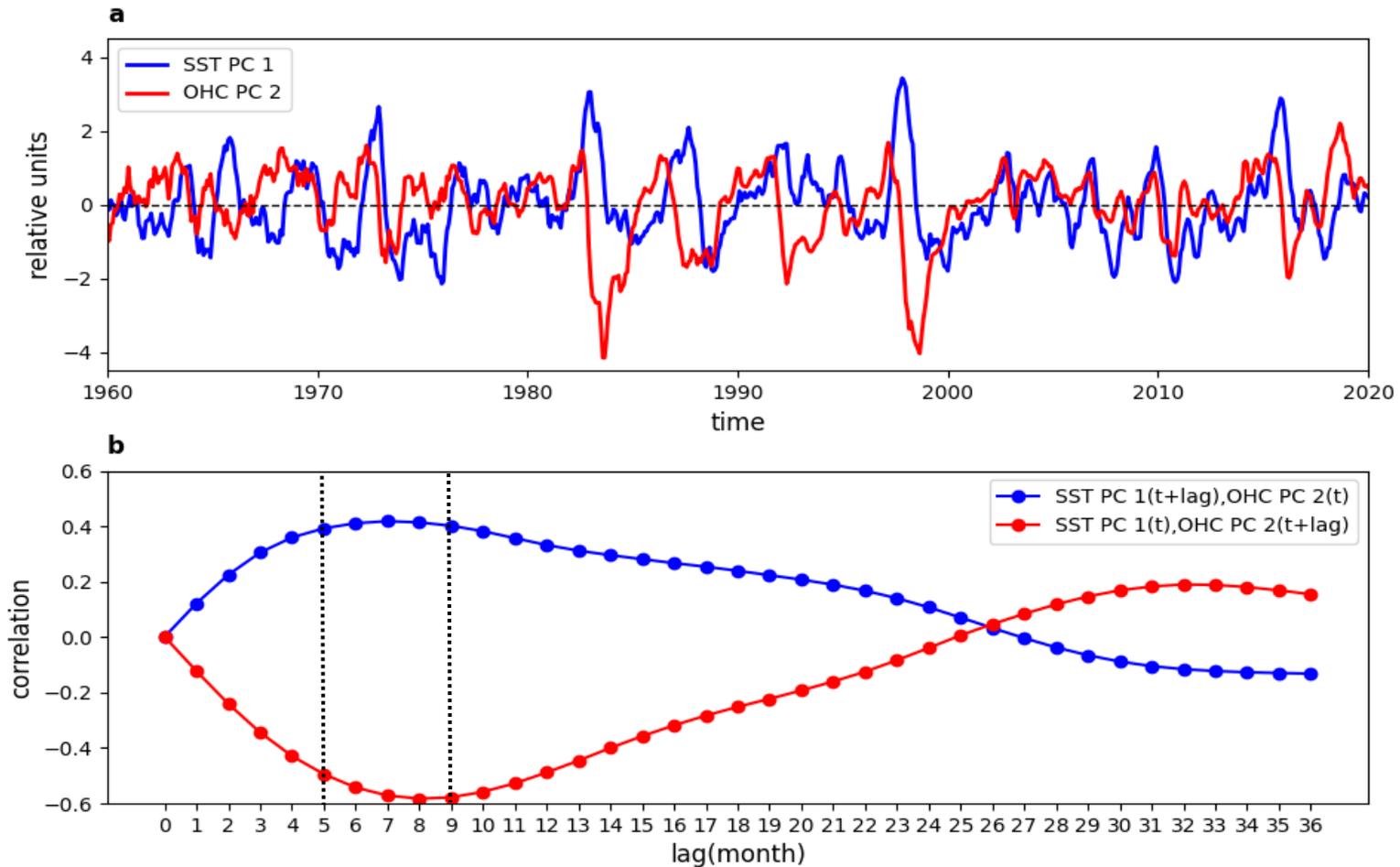
Зависимости между ведущими главными компонентами аномалий температуры поверхности океана и теплосодержания океана:



(a) ГК-1 и ГК-2 ТПО, (b) ГК-1 ТПО и ГК-1 ТО, (c) ГК-1 ТПО и ГК-2 ТО, (d) ГК-1 и ГК-2 ТО, (e) ГК-2 ТПО и ГК-1 ТО, (f) ГК-2 ТПО и ГК-2 ТО.

Цветные метки соответствуют декабрям тех лет, когда происходили экстремально сильные события Эль-Ниньо

Лаговые корреляции ГК-1 ТПО и ГК-2 ТО:



(a) Временные ряды ГК-1 ТПО (синяя линия) ГК-2 ТО (красная линия), нормированные на свои среднеквадратичные отклонения.

(b) Лаговые корреляции между этими временными рядами в случае, когда ТПО опережает ТО (синяя кривая) и наоборот (красная кривая)

Общий вид:

1. Линейные модели:

2. Нелинейные модели:

ДАННЫЕ (фазовые переменные):

Временные ряды ГК-1 среднемесячных аномалий ТПО и ГК-2 среднемесячных аномалий ТО с января 1960 года по декабрь 2020 года.

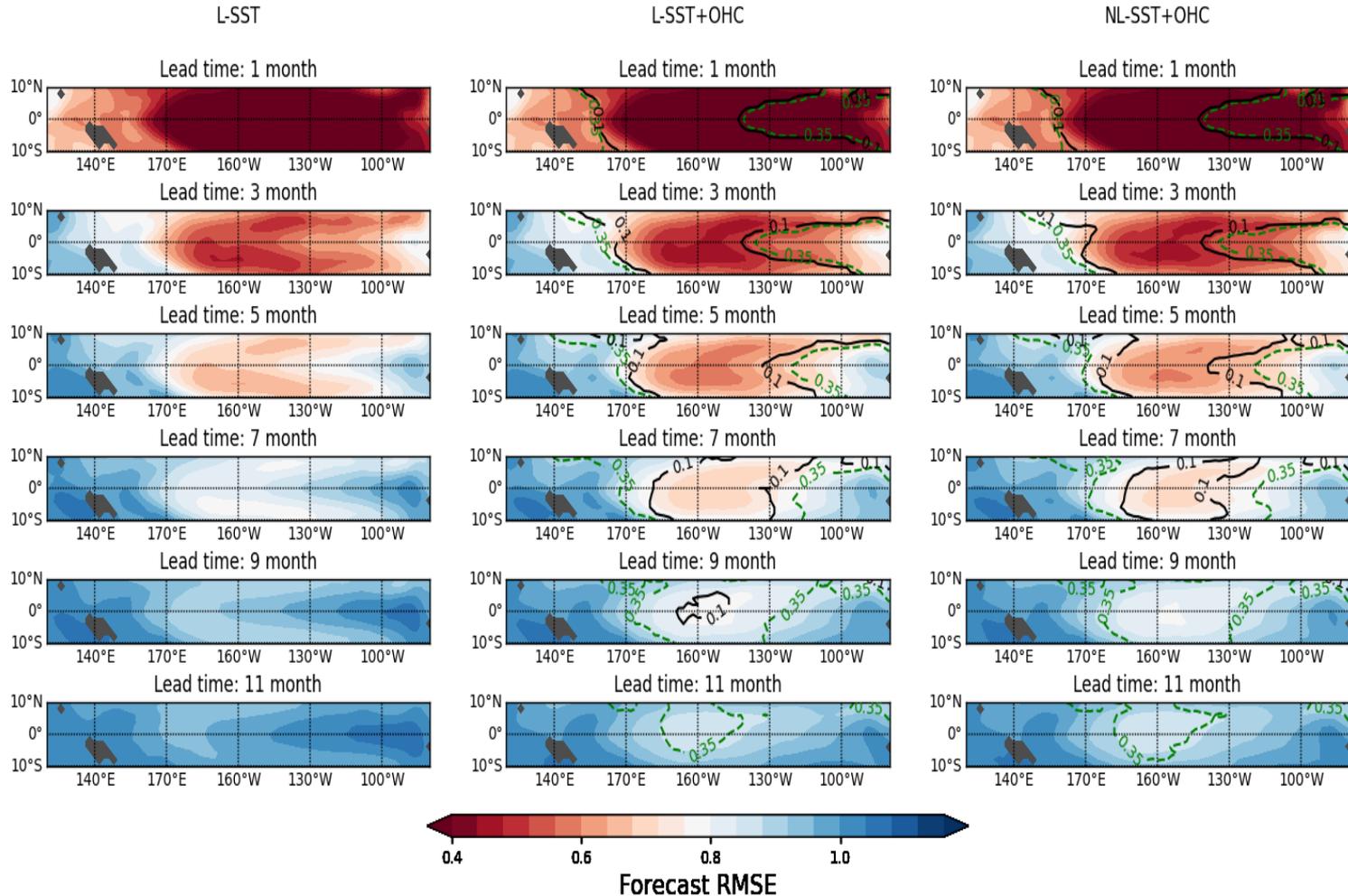
Обучающая выборка: с января 1960 года по декабрь 2007 года.

Контрольная выборка: с января 2008 года по декабрь 2020 года.

СРАВНИВАЛИСЬ ПРОГНОСТИЧЕСКАЯ СПОСОБНОСТЬ МОДЕЛЕЙ:

- **Независимые линейные** модели ТПО и ТО, построенные *отдельно* по временному ряду ГК-1 ТПО (модель L-ТПО) и временному ряду ГК-2 ТО (модель L-ТО);
- **Совместная линейная** модель (1.2)-(2.2) ТПО и ТО, построенная по временным рядам ГК-1 ТПО и ГК-2 ТО (модель L-ТПО+ТО);
- **Совместная нелинейная** модель (модель NL-ТПО+ТО)

Пространственные распределения КСКО ретроспективного прогноза нормированных аномалий ТПО для трех эмпирических моделей при различной дальности прогноза



Оптимальная ($l = 2, q = 1$)
линейная модель, основанная
только на данных ТПО

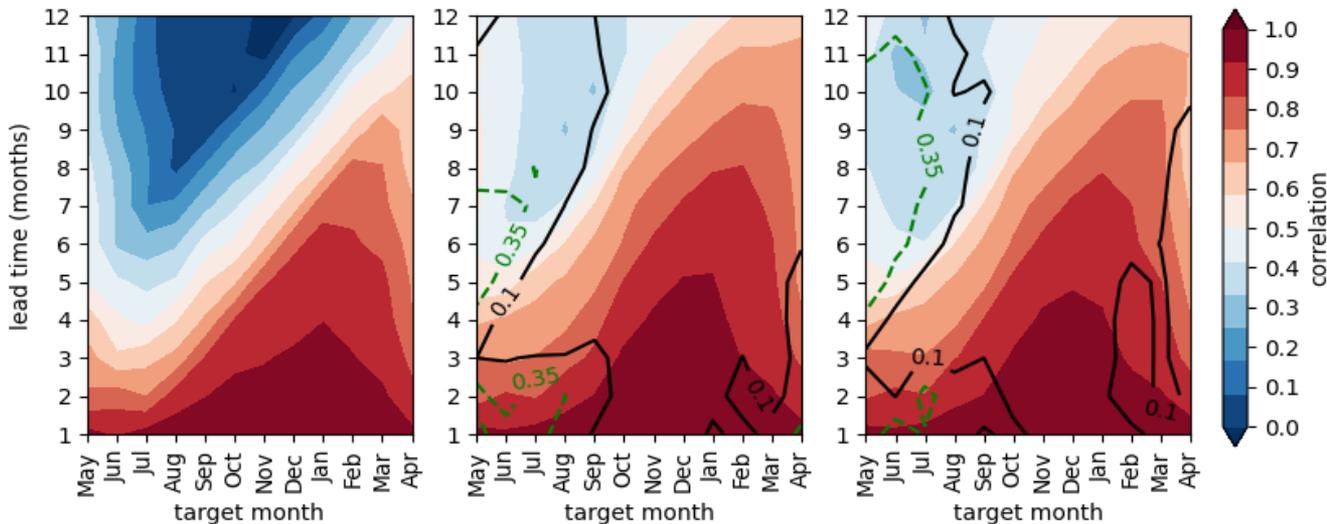
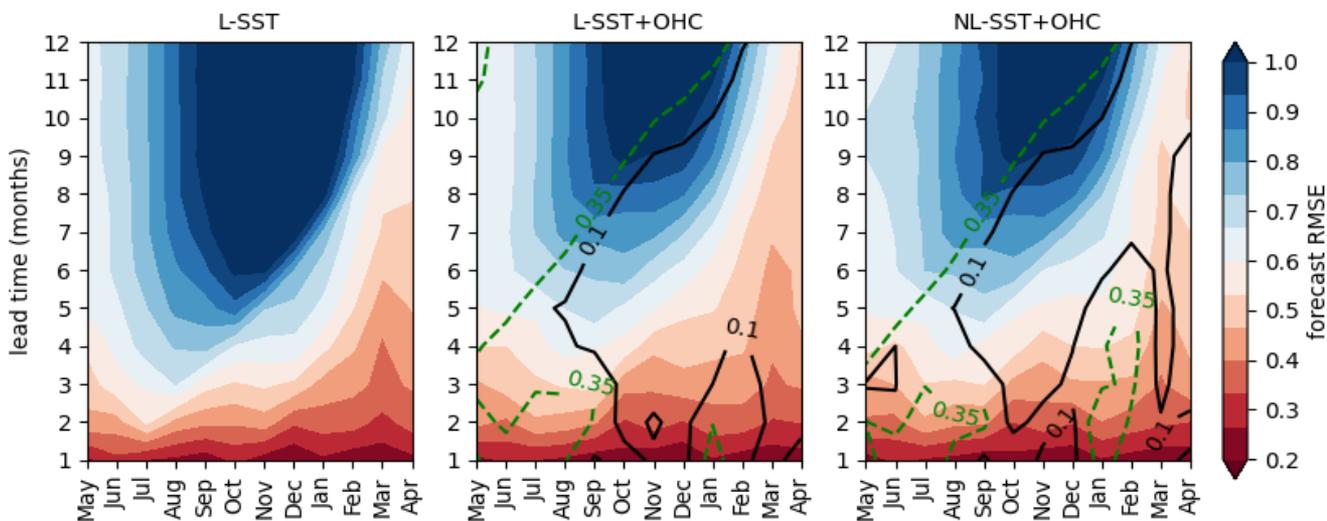
Оптимальная ($l = 2, q = 1$) совместная
линейная модель, основанная на данных
ТПО и ТО

Оптимальная ($l = 2, m = 5$)
совместная нелинейная модель



СЕЗОННАЯ ЗАВИСИМОСТЬ ПРЕДСКАЗАТЕЛЬНОЙ СПОСОБНОСТИ МОДЕЛЕЙ для индекса NINO 3.4

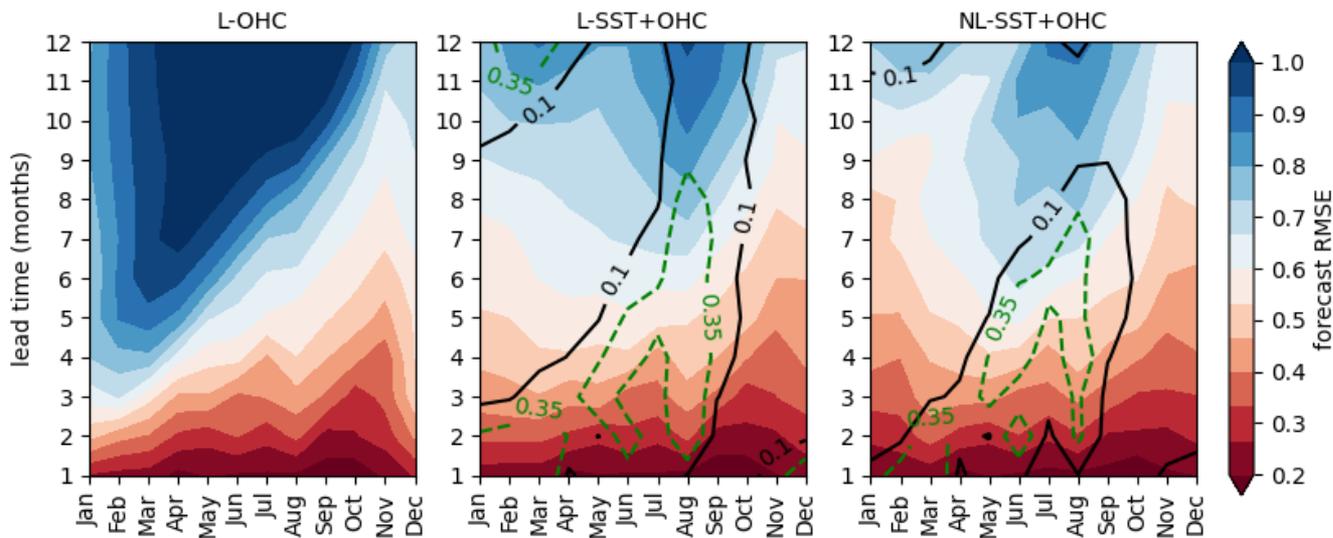
заблаговременность прогноза



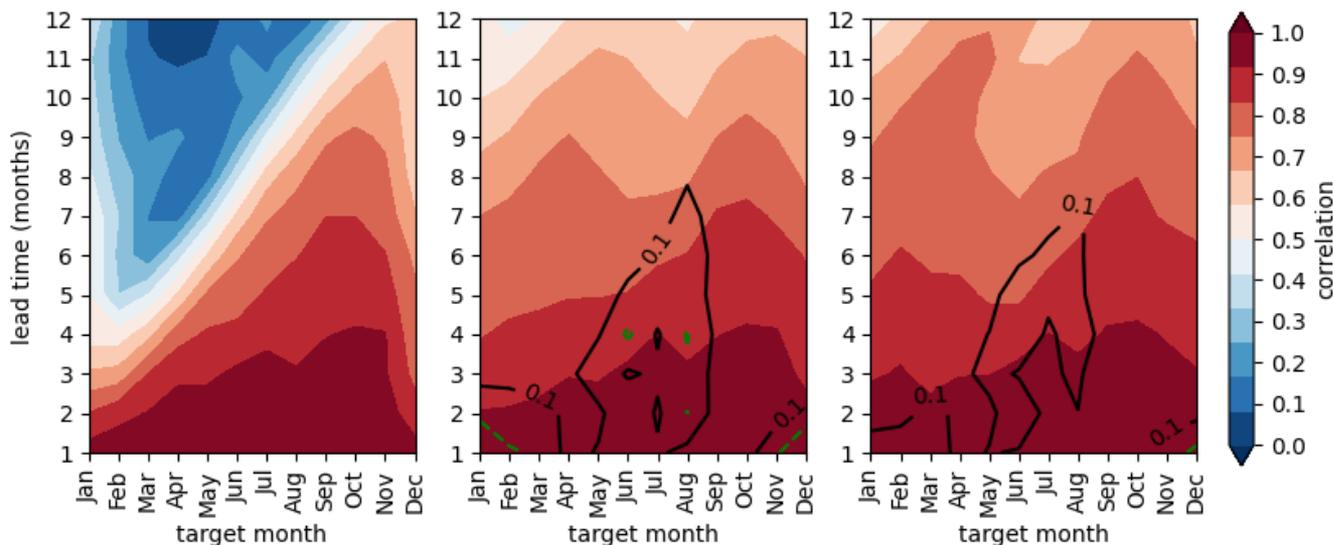
прогнозируемый месяц

СЕЗОННАЯ ЗАВИСИМОСТЬ ПРЕДСКАЗАТЕЛЬНОЙ СПОСОБНОСТИ МОДЕЛЕЙ для индекса аномалий теплосодержания (ГК-2 ТО)

заблаговременность прогноза



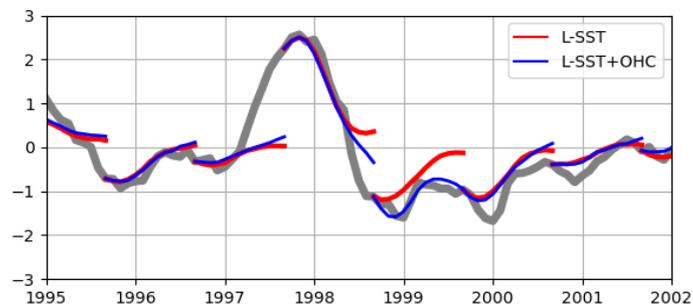
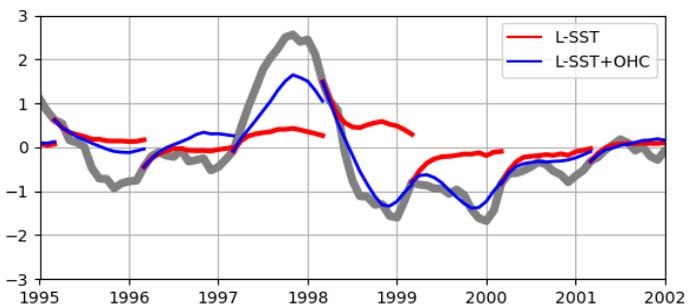
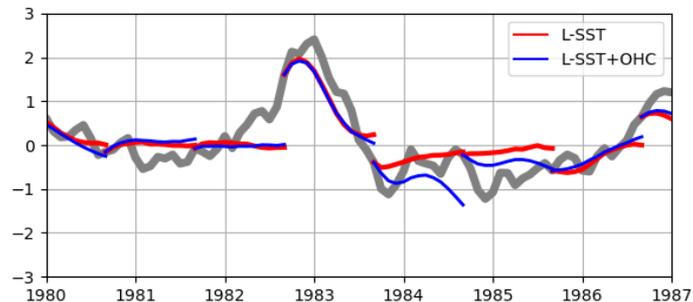
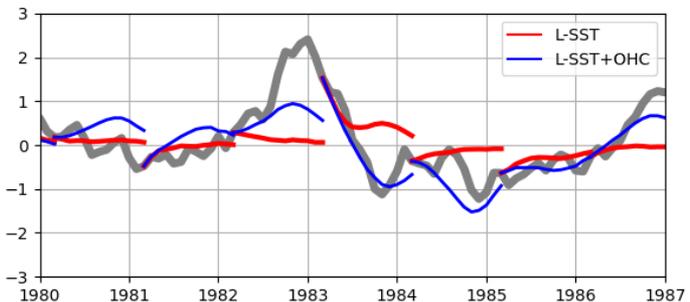
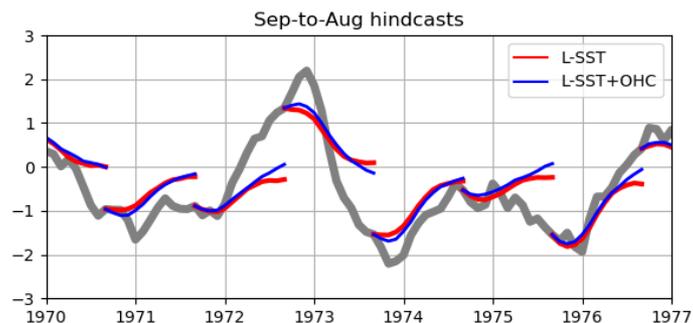
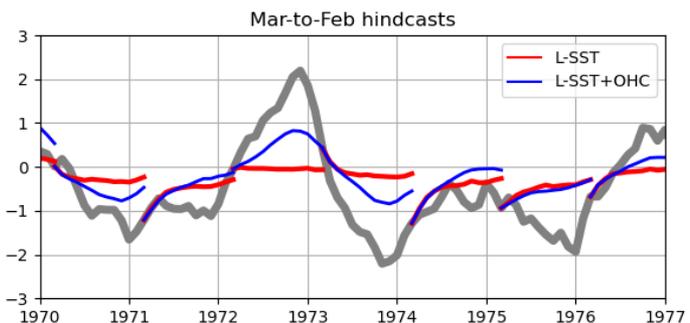
КСКО



коэффициент
корреляции

прогнозируемый месяц

ПРИМЕРЫ РЕТРОСПЕКТИВНЫХ (НА ОБУЧАЮЩЕЙ ВЫБОРКЕ) ПРОГНОЗОВ событий Эль-ниньо/Ла Нинья на основе индекса NINO 3.4: с помощью линейной модели, основанной только на данных ТПО (*красные кривые*) и линейной совместной модели (*синие кривые*)



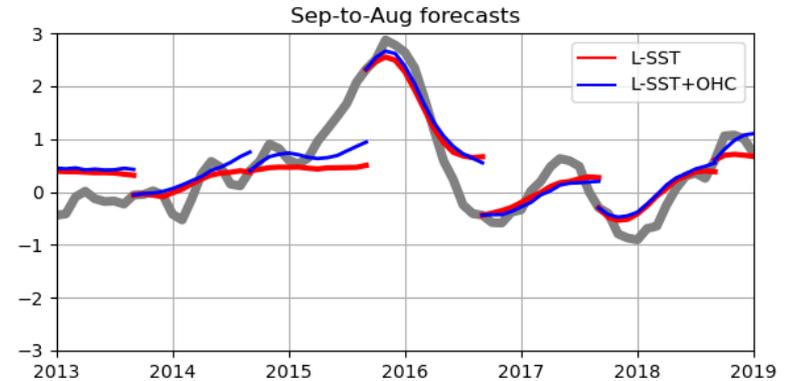
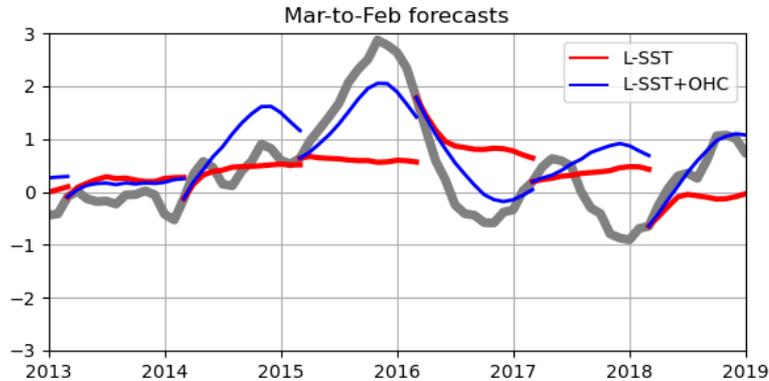
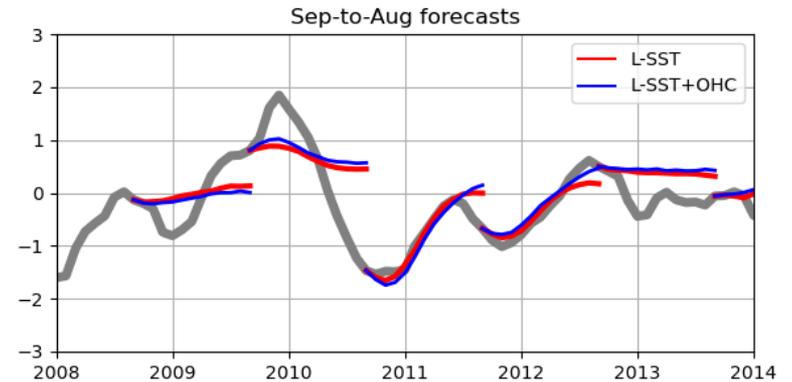
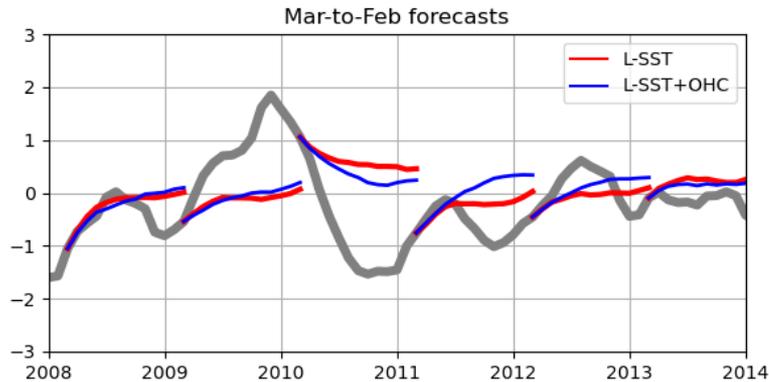
Прогнозы, начинающиеся с марта

Прогнозы, начинающиеся с сентября

Временной ряд индекса NINO 3.4 показан серым цветом

ПРИМЕРЫ РЕАЛЬНЫХ (НА ТЕСТОВОЙ ВЫБОРКЕ) ПРОГНОЗОВ СОБЫТИЙ ЭльНиньо/Ла-Ниньо на основе индекса NINO 3.4:

с помощью линейной модели, основанной только на данных ТПО (*красные кривые*) и линейной совместной модели (*синие кривые*)



Прогнозы, начинающиеся с марта

Прогнозы, начинающиеся с сентября

Временной ряд индекса NINO 3.4 показан серым цветом

ЧТО УДАЛОСЬ:

- 1. Найден индикатор («индекс»), характеризующий процессы накопления и разгрузки подповерхностных теплых вод в центральной и западной части Тихого океана - ключевой составляющей цикла колебания Эль-Ниньо.**
- 2. Выявлен фазовый сдвиг индексов, характеризующих динамику поверхностной температуры тропического Тихого океана и теплосодержания океана в этой области, составляющий 5-9 месяцев.**
- 3. Установлено, что, в отличие от традиционных индексов Эль-Ниньо найденный индекс ТО характеризуется зимним барьером предсказуемости.**
- 4. Показано, что совместные модели, учитывающие найденные индексы как равноценные динамические переменные, имеют существенно лучшую предсказательную способность по сравнению с моделями, основанными только на ТПО/ТО и позволяют преодолеть/существенно снизить сезонные (весенний в индексе ТПО и зимний в индексе ТО) барьеры предсказуемости.**
- 5. Прогностические способности совместных линейной и нелинейной совместных моделей практически не отличаются, что может свидетельствовать о том, что учет нелинейности в эмпирических прогностических моделях ЭНЮК не играет роли на внутригодовых масштабах.**

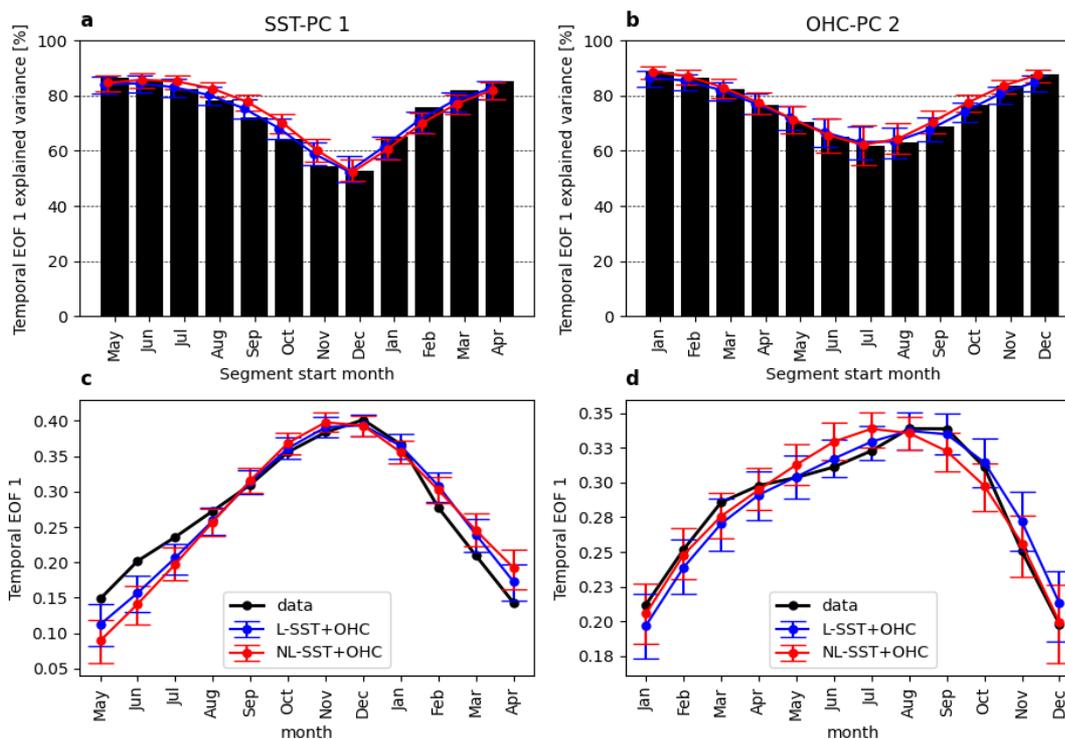
ВОЗНИКШИЕ ВОПРОСЫ:

Какова роль нелинейности в динамике ЭНЮК на больших временных масштабах?

2. Эмпирическое моделирование межгодовой изменчивости явления Эль-Ниньо

Выделение сезонных циклов по временным рядам ГК-1 ТПО и ГК-2 ТО на интервале с 1960 по 2020 год:

ЭОФ декомпозиция данных в 12-ти мерном пространстве среднемесячных данных



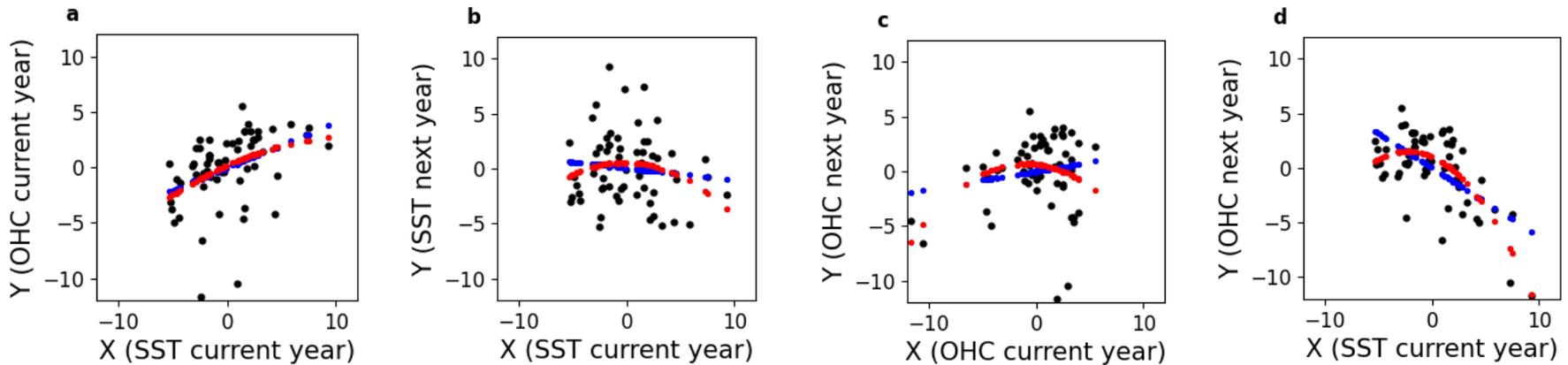
Верхняя панель: доли вариации ГК-1 ТПО (а) и ГК-2 ТО (b), описываемые главной временной ЭОФ в зависимости от начального месяца.

Нижняя панель: ЭОФ паттерны ГК-1 ТПО (c) и ГК-2 ТО (d).

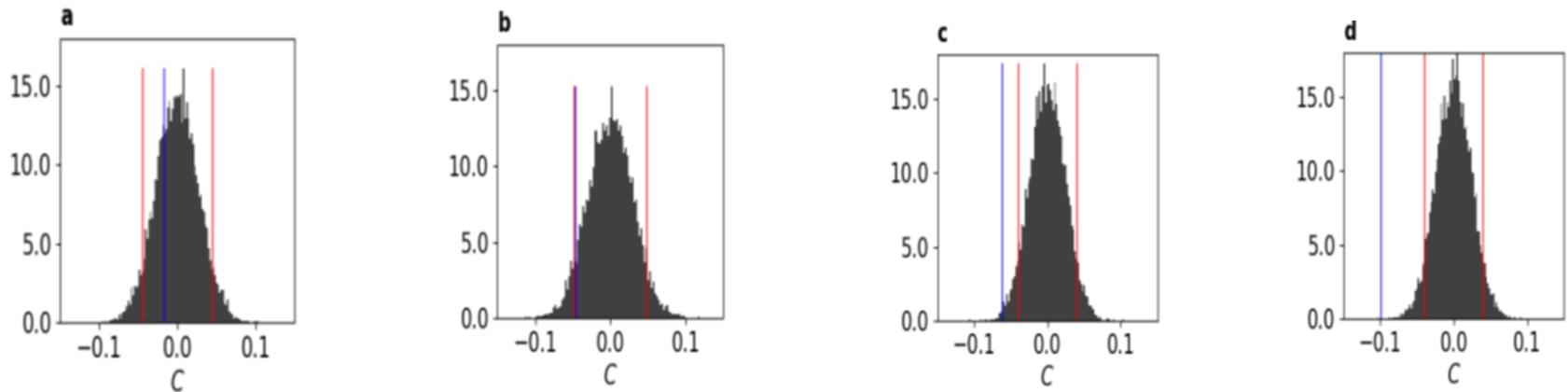
Черным: характеристики, полученные по данным.

Синим и красным: средние значения характеристик и их 90% доверительные интервалы, полученные по ансамблям из 1000 реализаций моделей L-ТПО+ТО и NL-ТПО+ТО

МЕЖГОДОВЫЕ ЗАВИСИМОСТИ АМПЛИТУД СЕЗОННЫХ ЦИКЛОВ ТПО И ТО



Зависимости между амплитудами сезонных циклов ТПО и ТО, полученные по данным (черные точки) и их аппроксимации на основе линейной (синяя кривая) и квадратичной (красная кривая) регрессионных моделей.



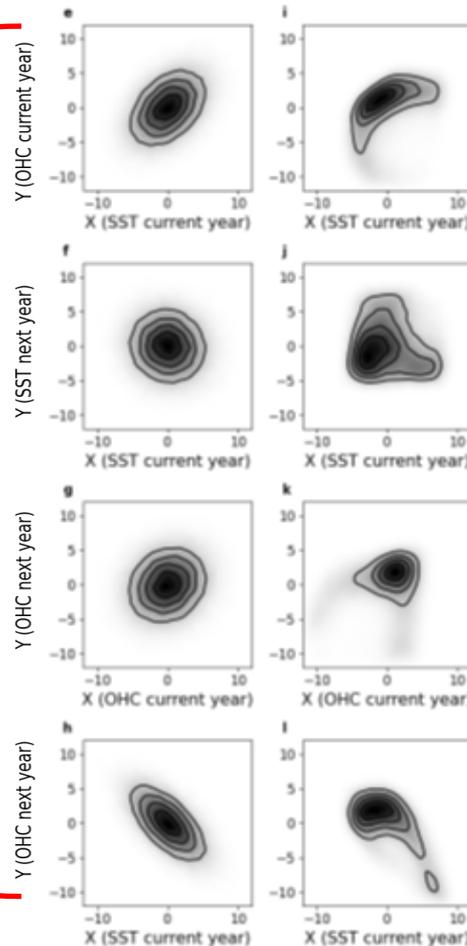
Проверка статистической значимости квадратичной аппроксимации зависимостей, показанных на верхней панели.

Синими линиями обозначены значения коэффициента, характеризующего нелинейности модели, обученной по данным.

Красными линиями отмечены 10-й и 90-й процентиля суррогатных распределений

Воспроизведение межгодовых зависимостей между компонентами ЭНЮК эмпирическими моделями

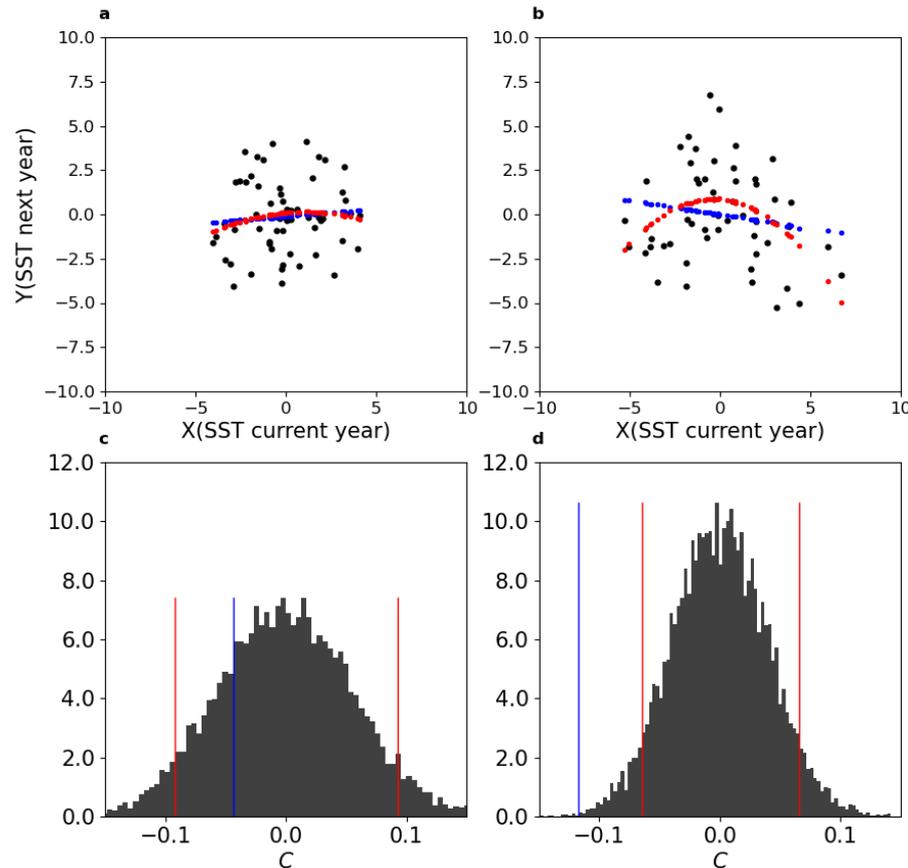
Квази-гауссовы ПВ
в случае линейной
эмпирической
модели



Негауссовы ПВ
в случае
нелинейной
эмпирической
модели

Плотности вероятности (ПВ) амплитуд сезонных циклов ТПО и ТО, рассчитанные по ансамблям из 1000 временных рядов, произведенным совместными (ТПО+ТО) моделями. Слева ПВ, построенная на основе линейной модели L- ТПО+ТО, справа – на основе нелинейной модели NL- ТПО+ТО

Изменение межгодовых зависимостей амплитуд годовых циклов аномалий ТПО с 1900 по 2014 год



Верхняя панель: зависимости между амплитудами годовых циклов аномалий ТПО на интервале с 1900 по 1954 год (a) и с 1960 по 2014 год (b).

Нижняя панель: проверка статистической значимости квадратичной аппроксимации ЭТИХ зависимостей.

ЧТО УСТАНОВЛЕНО:

Межгодовая зависимость амплитуд сезонных циклов аномалий ТПО в период с 1900 по 2014 год существенно изменилась: линейная зависимость в период 1900 по 1954 сменилась нелинейной в период с 1960 по 2014 год.

ВОЗМОЖНАЯ ПРИЧИНА :

**Рост числа экстремально сильных событий Эль-Ниньо во второй половине 20 века.
Влияние глобального потепления?**

ОБЩИЕ ВЫВОДЫ

ДИНАМИЧЕСКАЯ ЭМПИРИЧЕСКАЯ РЕКОНСТРУКЦИЯ НА ОСНОВЕ БАЙЕСОВА ПОДХОДА :

- **делает возможным исследование структуры фазового пространства системы, оператор эволюции которой неизвестен**
- **позволяет получить информацию о механизмах, определяющих динамику исследуемой системы**
 - **является инструментом для прогноза эволюции системы**
 - **дает уникальную возможность валидации существующих моделей, построенных на основе первых принципов**

СПАСИБО!